Channel Capacity Gain in Entanglement-Assisted Communication Protocols Based Exclusevly on Linear Optics and Single Photon Inputs

P. Lougovski[1, *] and D. B. Uskov[2, 3, †]

[1] *Quantum Information Science Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831*
[2] *Department of Mathematics and Natural Sciences, Brescia University, Owensboro, KY 42301*
[3] *Department of Physics and Engineering Physics, Tulane University, New Orleans, LA 70118*

Entanglement can effectively increase communication channel capacity as evidenced by dense coding that predicts a capacity gain of 1 *bit* when compared to entanglement-free protocols. However, dense coding relies on Bell states and when implemented using photons the capacity gain is bounded by 0.585 *bits* due to one's inability to discriminate between the four optically encoded Bell states. In this paper we study the following question: Are there alternative entanglement-assisted protocols that rely only on linear optics, coincidence photon counting and separable single photon input states and at the same time provide a greater capacity gain than 0.585 *bits*. We show that besides the Bell states there is a class of bipartite four-mode two-photon entangled states that facilitate an increase in channel capacity. We also discuss how the proposed scheme can be generalized to the case of two-photon N-mode entangled states for $N = 6, 8$.

PACS numbers: 03.67.Hk, 03.65.Ud

I. INTRODUCTION

Dense coding is a quantum communication technique that allows one to send two bits of classical information per single qubit transmitted over a quantum channel. This becomes possible when a sender and a receiver preshare a maximally entangled two-qubit state (Bell state). Of course, if no entanglement is shared, at most one bit of classical information can be communicated by sending a single qubit. Therefore, using entanglement as a communication resource increases channel capacity by 1 *bit*. The original protocol [1] – closely followed in all optical qubit implementations [2–5] – was designed for generic qubits and has two main requirements. First, the sender must be able to generate all four Bell states from any given Bell state via a local (single-qubit) operations only. Second, the receiver must be able to perform an unambiguous Bell state discrimination (typically using a two-qubit CNOT gate). For optical qubits – most suitable for a long distance communication – the first requirement can be readily fulfilled by combining spontaneous parametric down conversion entangled photon sources with linear optical devices such as beam splitters, polarization rotators and phase shifters. However, the second requirement of the original protocol is very stringent. Unfortunately, one cannot deterministically distinguish all four Bell states only by means of linear optical devices and coincidence measurements [6]. Either hyper entanglement [3–5, 7] or additional entangled ancillae are needed [8, 9] making all-optical implementations challenging. As a result, the entanglement-assisted channel capacity gain is bounded by $0.585 < 1$ *bits*. But even achieving the reduced bound is fairly difficult due to experimental imperfections. For instance, the actual channel capacity gain reported in [2] was ≈ 0.13 *bits*.

This prompted us to ask the following question: Are there alternative entanglement-assisted protocols (not based on Bell states) that utilize only linear optics, coincidence photon counting and input single photon product states and, at the same time, provide a greater capacity gain than 0.585 *bits*? We answer this question by constructing a communication protocol that satisfies the above resource constraints. In particular, we find an equivalence class of four-mode two-photon entangled states that can be prepared using product single photon input states with linear optical elements and transformed into each other by means of linear optical operations on (two) "local" modes of one of the parties and, at the same time, can be discriminated by a photon coincidence measurement deterministically. We formalize the problem mathematically in Section II and show that the explicit structure of these states as well as the detailed experimental setup that implements our protocol can be obtained by maximizing the mutual information between the sender and the receiver over all physical input states, local/global operations and detection schemes. By solving the optimization task numerically in Section III, we find that the communication channel capacity of our protocol with ideal detectors is 2 *bits*.

To verify that entanglement does indeed provide a gain in the channel capacity we determine the upper bound on the channel capacity with respect to all possible resource-equivalent entanglement-free protocols. We show in Section III that, under the condition of no vacuum detection, no entanglement-free protocol can achieve the channel capacity greater than 1 *bit* and, thus, our protocol allows the sender to communicate an extra 1 *bit* of classical information which is better than the gain of 0.585 *bits* offered by dense coding with linear optics.

On the other hand, we demonstrate that if vacuum detection is allowed but detectors are imperfect then our

[*] Electronic address: lougovskip@ornl.gov
[†] Electronic address: dmitry.uskov@brescia.edu

protocol still provides a detection-efficiency-dependent channel capacity gain. For example, the gain is ≈ 0.27 *bits* for the state-of-the-art superconducting single photon detectors. Also, our protocol can be extended to the case of two photons shared among $N > 4$ modes and in the Section IV we consider two scenarios with $N = 6$ and $N = 8$. We show that in both cases solutions exist that provide a capacity gain over corresponding entanglement-free protocols.

II. CHANNEL CAPACITY FORMALISM FOR FOUR-MODE TWO-PHOTON COMMUNICATION PROTOCOLS

When implementing an abstract two-qubit system using single photons two so-called "dual rail" schemes are prevalent. The first one is the polarization encoding where the logical zero and one states of each qubit are realized as the horizontally $|H\rangle$ and vertically $|V\rangle$ polarized single photon in a given spatial mode and the second is the spatial mode encoding where a single photon placed in either one of two spatial modes i.e. $|0,1\rangle$ and $|1,0\rangle$ represents the logical zero and one states. The schemes can be mapped onto each other by setting $|H\rangle \equiv |0,1\rangle, |V\rangle \equiv |1,0\rangle$. An arbitrary local (involving modes of one of the parties only) operation can be performed by means of linear optical elements such as a polarization rotator or a beam splitter in the case of dual rail encoding. This class of operations will map the two-qubit computational space $\mathbb{C}^4 = span\{|0,1,0,1\rangle, |0,1,1,0\rangle, |1,0,0,1\rangle, |1,0,1,0\rangle\}$ onto itself. However, when implementing an operation involving spatial modes from both parties with linear optics an input state from $\mathbb{C}^4$ may end up in a larger Hilbert space $\mathbb{C}^{10}$. For example, consider the action of a 50/50 beam splitter between modes 2 and 3 on the state $|\psi\rangle = |0,1\rangle \otimes |1,0\rangle \in \mathbb{C}^4$. The resulting state $|\tilde{\psi}\rangle = \frac{1}{\sqrt{2}}(|0,0,2,0\rangle - |0,2,0,0\rangle)$ actually lies outside of $\mathbb{C}^4$ in a larger Hilbert space $\mathbb{C}^{10}$. In fact, this becomes an issue when one tries to implement a quantum computer using only linear optical transformations. But for quantum communication problems it may be more advantageous to operate in the full (two photons in four modes) Hilbert space $\mathbb{C}^{10}$.

Indeed, consider the following communication protocol implemented using $N = 4$ modes and $n = 2$ photons (see Fig. 1). The source generates a special initial two-photon four-mode state $|\Psi_{in}\rangle \in \mathbb{C}^{10}$ and sends half of the modes to Alice and the other half to Bob. For the sake of concreteness we assume that Alice gets modes labeled 1 and 2 (1 through N_A if N modes are shared). As a result, Alice and Bob share the state $|\Psi_{in}\rangle$. Upon receiving her part of the state, Alice transforms her modes using one of the four (or potentially more) predetermined *two-mode unitary* transformations $U_i, i = 1, \cdots, 4$. She chooses which unitary operation U_i to apply according to a probability distribution $p(U_i)$ and does not disclose her choice to Bob. Next, Alice sends her part of the state to Bob who now has one of four possible two-photon four-mode states $|\psi_i\rangle$ with probability $p(|\psi_i\rangle) = p(U_i)$. Bob wants to learn which unitary transformation U_i was performed by Alice i.e. which state $|\psi_i\rangle$ he has at hand. To do that he sends all four modes through a detection setup that performs a *four-mode unitary* transformation U_{Bob} and measures the projection of the output state onto the two-photon four-mode Fock basis states $\{|\phi_1\rangle = |2000\rangle, |\phi_2\rangle = |1100\rangle, \cdots, |\phi_9\rangle = |0011\rangle, |\phi_{10}\rangle = |0002\rangle\}$. Using the outcome of the measurement Bob tries to guess the index i of the state that Alice has sent.

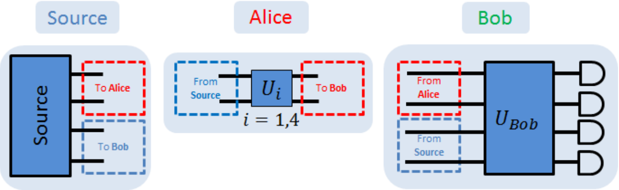

FIG. 1: (Color on-line) Illustration of the proposed entanglement-assisted communication protocol. Alice and Bob share a two-photon four-mode state $|\Psi_{in}\rangle$ distributed by the source. Alice performs a "local" mode transformation U_i on her two modes and sends them to Bob who performs a coincidence detection on all four modes and estimates index i.

In qubit-based communication protocols Alice receives a single qubit (i.e. photon) from the source. The photon, depending on the encoding scheme, is distributed over two (polarization or spatial) modes. The local unitary operations $U_i, i = 1, \cdots, 4$ that Alice performs on her qubit do not change the total number of photons she communicates to Bob. This provides an implicit constraint on the average number of photons used to communicate a single character i.e.

$$\sum_{i=1}^{4} p_i \langle\psi_i| a^\dagger_{A1} a_{A1} + a^\dagger_{A2} a_{A2} |\psi_i\rangle = 1, \tag{1}$$

where a_{A1} and a_{A2} are photon annihilation operators for Alice's mode 1(2) and p_i is the probability that Alice performs an operation U_i. In our qubit-less protocol the source can potentially provide a state $|\Psi_{in}\rangle$ that violates the constraint in Eq.(1). To make our protocol comparable to dense coding with linear optics in terms of photonic resources needed to communicate a single character we, therefore, explicitly require that the state provided by the source to Alice and Bob is such that it satisfies the photon number constraint in Eq.(1).

Naturally, the best case scenario in terms of information transmission(assuming a noiseless quantum channel and perfect detectors) for our protocol is when Alice can prepare four orthogonal states $|\psi_i\rangle$ subjected to constraints Eq.(1) and Bob can unambiguously detect which state $|\psi_i\rangle$ Alice has sent him just by means of photon coincidence measurements. In this case, analogously to the original dense coding proposal [1], Alice and Bob can communicate two bits of information (provided $p_i = \frac{1}{4}\forall i$) by sending on average one photon in just two modes instead of four. However, whether such states $|\psi_i\rangle$ (equivalently a state $|\Psi_{in}\rangle$ and two-mode unitaries U_i) exist is an open question. Also, to conclude that our protocol is indeed of a dense coding type, we need to determine the largest amount of classical information that Alice and Bob can share by utilizing at most two single photons and two modes under the constraint Eq.(1). Only if the latter is less than 2 *bits* our protocol demonstrates an information gain. To answer these questions we need to formulate the problem in terms of communication channel capacity determination.

From the information-theoretical perspective Alice and Bob form a two-mode two-photon communication system with two auxiliary modes in which Alice encodes each letter of her message using a two-bit alphabet $\mathcal{X} = \{|\psi_1\rangle, \cdots, |\psi_4\rangle\}$ with probability $p(\psi_i) = p(U_i)$. The receiver, Bob, detects the message sent by Alice as a collection of random signals from the set $\mathcal{Y} = \{|\phi_1\rangle, \cdots, |\phi_{10}\rangle\}$ with the conditional probability $p(\mathcal{Y}|\mathcal{X}) = p(\psi_i|\phi_j) = |\langle\phi_i|\psi_j\rangle|^2, i = 1, \cdots, 10; j = 1, \cdots, 4$ and applies a decoding rule to estimate the original message. Note that $p(\mathcal{Y}|\mathcal{X})$ is a function of the initial state $|\Psi_{in}\rangle$, Alice's unitary operations $U_i, i = 1, \cdots, 4$ and Bob's detection setup U_{Bob}. In this context, finding the highest rate (in bits) at which information can be sent from Alice to Bob is equivalent to determining information channel capacity of the Alice-Bob system. The information channel capacity C of Alice-Bob channel is defined as [10],

$$C = \max I(\psi; \phi), \tag{2}$$

where the maximization is performed over all possible input distributions $p(\psi_i)$, states $|\Psi_{in}\rangle$, unitary mode transformations U_i, U_{Bob} and is subjected to the constraint in Eq.(1); I is the mutual information,

$$I(\psi; \phi) = \sum_{j=1}^{4}\sum_{k=1}^{10} p(\psi_j, \phi_k) \log \frac{p(\psi_j|\phi_k)}{p(\psi_j)}. \tag{3}$$

Here $p(\psi_j, \phi_k) = p(\psi_j|\phi_k)p(\phi_k)$ denotes the joint probability of Alice preparing the state $|\psi_j\rangle$ and Bob detecting the state $|\phi_k\rangle$. The marginal probability $p(\phi_k)$ is defined as $p(\phi_k) = \sum_{\psi_j} p(\psi_j, \phi_k)$.

By definition, the mutual information $I(\psi; \phi)$ is a concave function of $p(\psi_i)$ (3 independent real-valued parameters) and a non-concave function of the unitary mode transformation matrices $U_i \in U(2), i = 1, \cdots, 4$ (16 real-valued parameters), and $U_{Bob} \in SU(4)$ (15 independent real-valued parameters) and the input state $|\Psi_{in}\rangle$ that we parametrize using ten complex parameters c_k (18 independent real-valued parameters [12]) as $|\Psi_{in}\rangle = \sum_{k=1}^{10} c_k|\phi_k\rangle$. We can always set one of the matrices U_i to be the identity $\mathbb{1}$ matrix which will leave us with only three independent 2×2 unitary matrices. Therefore, the total number of real-valued optimization parameters in Eq.(2) is 48.

	Dense Coding	Proposed Protocol
Entanglement	Bell States	Multi-mode Entanglement
# of photons sent by Alice	1	1 on average
Total # of photons	2	2
Total # of modes	4	4
Operations	Linear Optics	Linear Optics
Detection	Coincidence	Coincidence

TABLE I: Resource overview for the proposed entanglement-assisted protocol and conventional dense coding with photonic qubits

Since we set Alice's alphabet to only four letters, the global maximum of the channel capacity C over all possible physical setup parameters cannot in principle exceed $\log_2(4) = 2$ bits. This follows form the definition of the mutual information in Eq.(3). We observe that for any two random variables $\mathcal{X}$ and $\mathcal{Y}$: $I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y})$, where H denotes Shannon entropy [10]. Since $H \geq 0$, the maximum of $I(\mathcal{X}; \mathcal{Y})$ is achieved when $H(\mathcal{X})$ is maximal ($= \log_2(|\mathcal{X}|)$) and $H(\mathcal{X}|\mathcal{Y})$ is minimal (=0) i.e. $\max I(\mathcal{X}; \mathcal{Y}) = \log_2(|\mathcal{X}|)$. However, it is not clear if this bound is physically attainable. Also, due to the non-concave nature of the optimization objective function many local maximums may exist. Of course, when optimizing $I(\psi; \phi)$ numerically we are interested in finding a supremum of all local maximums and hope that it is 2 bits. Note that because I is concave in parameters $p(\psi_i)$, if the global (2 bits) maximum is attained, using the preceding argument one can immediately show that the only possible values of $p(\psi_i) = \frac{1}{4}\forall i$. Therefore, we can further reduce the number of real optimization parameters to 45 by setting $p(\psi_i) = \frac{1}{4}$. We conclude this section by providing an overview of the resource requirements for our entanglement-assisted communication protocol as well as for the photon-based dense coding protocol in Table I. The only difference between the protocols comes from the number of photons that Alice communicates to Bob. In our protocol this number is one on average. Whereas in the dense coding protocol Alice always sends Bob one photon. This requirement also restricts a class of resource-equivalent entanglement-free protocols to those that use one photon on average.

III. FOUR-MODE TWO-PHOTON PROTOCOL ANALYSIS

A. Optimization Results

First, to test our approach, we solved the optimization problem in Eq.(2) using a fixed state $|\Psi_{in}\rangle$ provided by the source. We set $|\Psi_{in}\rangle$ to be equal to one of the Bell states (it does not matter which Bell state is selected, optimization works equally well for all of them) and found by numerical optimization that in this case $C = \log_2 3$. Moreover, Alice's mode transformation matrices that correspond to this solution are the same as the ones originally proposed by Bennett and Wiesner [1]. It means that by setting the initial state to a Bell state the conventional Bell state-based dense coding protocol [1, 2] is recovered.

Next, we have discovered, by using gradient-based optimization methods, that the global maximum ($C = 2$ *bits*) is indeed achievable in $\mathbb{C}^{10}$. The structure of globally optimal solutions encountered in our numerical search can be parametrized as follows. All globally optimal input states $|\Psi_{in}\rangle$ prepared by the source are, up to a swap of any two modes, equivalent to the state,

$$|\Psi_{in}\rangle = \frac{1}{2}(|1,1,0,0\rangle + |0,1,1,0\rangle + |1,0,0,1\rangle + |0,0,1,1\rangle). \tag{4}$$

For example, the following input state

$$|\tilde{\Psi}\rangle = \frac{1}{2}(|1,0,1,0\rangle + |0,1,1,0\rangle + |1,0,0,1\rangle + |0,1,0,1\rangle), \tag{5}$$

obtained from $|\Psi_{in}\rangle$ by swapping modes 2 and 3 also leads to the globally optimal solution with $C = 2$ *bits*.

Moreover, $|\Psi_{in}\rangle$ in Eq.(4) also defines a class of globally optimal input states that are equivalent to $|\Psi_{in}\rangle$ up to a four-mode unitary transformation:

$$U_t = \begin{bmatrix} U_A & 0 \\ 0 & U_B \end{bmatrix}, \tag{6}$$

where $U_{A,B}$ are arbitrary unitary matrices $\in U(2)$,

$$U_A = \begin{bmatrix} e^{i\phi_1}\cos\theta_1 & -e^{i\phi_2}\sin\theta_1 \\ e^{i\phi_3}\sin\theta_1 & e^{i(\phi_2+\phi_3-\phi_1)}\cos\theta_1 \end{bmatrix}, \tag{7}$$

$$U_B = \begin{bmatrix} e^{i\phi_4}\cos\theta_2 & -e^{i\phi_5}\sin\theta_2 \\ e^{i\phi_6}\sin\theta_2 & e^{i(\phi_5+\phi_6-\phi_4)}\cos\theta_2 \end{bmatrix}, \tag{8}$$

and parameters $\theta_{1,2}$, $\phi_{1,\cdots,6}$ are arbitrary angles $\in [0, 2\pi]$.

Given matrices U_A and U_B, Alice's globally optimal mode transformation matrices (acting on modes 1 and 2) can be decomposed as $U_1 = U_A^{-1}U_C$, $U_2 = -U_A^{-1}\sigma_z U_C$, $U_3 = -U_2$, $U_4 = U_2 \cdot U_3$, where U_A is defined in Eq.(7), U_C is an arbitrary 2×2 unitary matrix $\in U(2)$ with a similar parametrization and σ_z denoted Pauli sigma Z matrix.

Lastly, Bob's four-mode transformation matrix U_{Bob} can be represented as follows,

$$U_{Bob} = \frac{1}{\sqrt{2}} \begin{bmatrix} U_C^{-1} & 0 \\ 0 & U_B^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \tag{9}$$

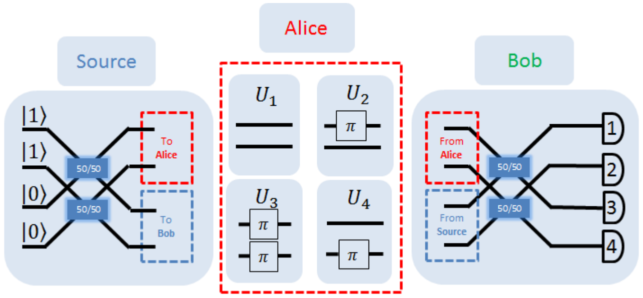

FIG. 2: (Color on-line) Linear optical implementation of the proposed entanglement-assisted protocol. Blue rectangles represent 50/50 beam splitters and transparent squares represent 180° phase shifters

Implementing this mode transformation matrix in an experiment, Bob will detect four distinct coincidence patterns: a coincidence between detectors in modes 1 and 2 correspond to Alice's choice U_1, modes 2 and 3 correspond to U_2, modes 3 and 4 correspond to U_3, modes 1 and 4 correspond to U_4. We remark that this outcome mapping is not unique. Other choices of coincidence assignment are possible and can be realized by additional four-mode unitary rotation on Bob's end i.e. $U_{Bob} \to U_{Bob} \cdot U_{swap}$.

We emphasize that U_t, $U_1, \cdots, U_4$ and U_{Bob} define globally optimal unitary *mode* transformation. The result of their action on the input state $|\Psi_{in}\rangle$ can be determined in the following fashion. Let us denote input mode photon creation operators as $a_k^\dagger, k = 1, \cdots, 4$, then,

$$|\Psi_{in}\rangle = \frac{1}{2}(a_1^\dagger a_2^\dagger + a_2^\dagger a_3^\dagger + a_1^\dagger a_4^\dagger + a_3^\dagger a_4^\dagger)|0\rangle, \tag{10}$$

where $|0\rangle$ is a four-mode vacuum state. Consider first the unitary mode operation U_t defined earlier. It acts on the creation operators $a_k^\dagger$ as a linear transform i.e. $a_k^\dagger \to \sum_{j=1}^{4}(U_t)_{kj}a_j^\dagger$, where $(U_t)_{kj}$ are the matrix elements of U_t. As a result, the input state $|\Psi_{in}\rangle$ in Eq.(10) is transformed to

$$|\Psi_t\rangle = \frac{1}{2}(\sum_{i,j=1}^{4}[(U_t)_{1i}(U_t)_{2j} + (U_t)_{2i}(U_t)_{3j} + (U_t)_{1i}(U_t)_{4j} + (U_t)_{3i}(U_t)_{4j}]a_i^\dagger a_j^\dagger)|0\rangle. \tag{11}$$

Similarly, unitary mode transformations $U_1, \cdots, U_4$ and U_{Bob} can now be applied to the state $|\Psi_t\rangle$ in Eq.(11) in sequence, resulting in the desired output state at Bob's detectors.

The structure of the globally optimal solution is most transparent when $U_A = U_B = U_C = \mathbb{1}$. In this case the physical setup that implements our communication protocol is illustrated in Fig.(2). Surprisingly, its structure is equivalent to a double Mach-Zehnder interferometer.

The source prepares the state $|\Psi_{in}\rangle$ in Eq.(4) by placing a separable state containing two single photons in modes 1 and 2 i.e. $|1,1,0,0\rangle$ onto two 50/50 beam splitters coupling modes (1,3) and (2,4). To transform her two modes, Alice needs only 180° phase shifters. Note that the transformation U_3 applies a phase shift to both Alice's modes with respect to Bob's modes. Bob recombines modes (1,3) and (2,4) on two 50/50 beam splitters and sends them for the coincidence detection. In the case of noiseless channel and perfect detectors this converts the states $|\psi_1\rangle \rightarrow |1,1,0,0\rangle$, $|\psi_2\rangle \rightarrow |0,1,1,0\rangle$, $|\psi_3\rangle \rightarrow |0,0,1,1\rangle$, $|\psi_4\rangle \rightarrow |1,0,0,1\rangle$ which results in the channel capacity $C = 2\ bits$. Also note that photon number resolving detectors are not required by Bob since he relies on a coincidence detection.

B. Channel Capacity of Resource-Equivalent Entanglement-Free Protocols

Let us now calculate the channel capacity for the case when Alice and Bob do not share any entanglement in order to determine whether our entanglement-assisted protocol provides any advantage for information transmission purposes. For that we consider the following resource-equivalent scenario: Alice prepares a *two-mode* state $|\tilde{\psi_i}\rangle, i = 1, \cdots, M$, corresponding to a character from a M-letter alphabet $\mathcal{X}$ (M is to be determined later), with probability $p(\tilde{\psi_i})$ and sends it to Bob who can perform an arbitrary *two-mode* unitary transformation U_{Bob2} before he detects the received state using two photo detectors that are *not* number resolving (because the correspondent entanglement-based protocol does not require photon-number-resolving detectors).

To facilitate a fair comparison between the entanglement-free and entanglement-assisted protocols, we assume that Alice's resources per character are limited to at most *two single photons* and she may use only linear optical elements(i.e. only two-mode unitary operations) to prepare states $|\tilde{\psi_i}\rangle$. In this case, inputs to Alice's state preparation setup are limited to $|0,0\rangle, |0,1\rangle, |1,0\rangle, |1,1\rangle$. Linear optical operations that Alice may perform to prepare states $|\tilde{\psi_i}\rangle$ conserve the total number of photons and, thus, $|\tilde{\psi_i}\rangle$ cannot be a superposition of states with different total photon numbers such as, for instance, $|0,0\rangle + |1,1\rangle$ or $|0,0\rangle + |0,1\rangle$, etc. Also, the states $|\tilde{\psi_i}\rangle$ must be mutually orthogonal. If they are not, Bob will be forced to discriminate between non-orthogonal states since he can only apply a two-mode unitary transformation (unitary transformations do not change the overlap between the states $|\tilde{\psi_i}\rangle$) to all states he receives. This will effectively reduce the amount of information that Bob and Alice can exchange and make such a communication protocol suboptimal.

The subspaces that contain states with different total number of photons are orthogonal and we only need to determine how many orthogonal states can be created by Alice in each subspace by means of a two-mode unitary transform. The subspace spanned by vacuum $|0,0\rangle$ is invariant under two-mode unitary transforms and Alice can use $|0,0\rangle$ as $|\tilde{\psi_1}\rangle$. The one-photon subspace spanned by $|0,1\rangle$ and $|1,0\rangle$ trivially contains two orthogonal states e.g. $|\tilde{\psi}_{2,3}\rangle = \frac{1}{\sqrt{2}}(|0,1\rangle \pm |1,0\rangle)$ that can be generated using a 50/50 beamsplitter. Finally, it is straightforward to show that only one state orthogonal to $|1,1\rangle$ can be generated by two-mode unitary operations in the two-photon subspace. Hence, Alice can choose $|\tilde{\psi_4}\rangle = \frac{1}{\sqrt{2}}(|2,0\rangle - |0,2\rangle)$ and $|\tilde{\psi_5}\rangle = |1,1\rangle$. This implies that Alice may prepare up to $M = 5$ mutually orthogonal states by using linear optical elements.

We now can calculate the channel capacity for the above choice of the states $|\tilde{\psi_i}\rangle, i = 1, 5$, maximizing the mutual information function over Bob's unitary transformation U_{Bob2} and probabilities $p(|\tilde{\psi_i}\rangle)$, subjected to the average photon number constraint in Eq.(1) which for this case reads,

$$p_2 + p_3 + 2p_4 + 2p_5 = 1. \tag{12}$$

We found, by solving the optimization task numerically, that Bob's optimal transformation is a 50/50 beam splitter and the channel capacity in this case amounts to 2 *bits*. The optimal values for p_i's are $p_1 = p_2 = p_3 = p_5 = 0.25$; $p_4 = 0$. Therefore, Alice can equivalently (up to a two-mode unitary transformation) use the following set of states to encode her message:

$$\{|\tilde{\psi_i}\rangle\} = \{|0,0\rangle, |0,1\rangle, |1,0\rangle, |1,1\rangle\}. \tag{13}$$

This is because Bob's detectors are not number resolving, and he cannot discriminate a single photon event from a two-photon event. Therefore, even if Alice can prepare five mutually orthogonal states the amount of information that Bob can extract from them is the same as if Alice used the four state alphabet in Eq.(13).

At first sight one may conclude that if Bob has ideal(100% efficient) detectors and additional synchronization information to distinguish the vacuum signal $|0,0\rangle$ from no signal events, he could discriminate the states $|\tilde{\psi_i}\rangle$ in Eq.(13) perfectly which implies that the channel capacity of the entanglement-free protocol is also 2 *bits*. However, if the vacuum detection is ruled out, it would further reduce the set $\{|\tilde{\psi_i}\rangle\}$ to only three states i.e. $\{|\tilde{\psi_i}\rangle\} = \{|0,1\rangle, |1,0\rangle, |1,1\rangle\}$. Moreover, by combining Eq.(12) with the normalization condition $\sum_{i=1}^{5} p_i = 1$ and setting $p_1 = p_4 = 0$ we immediately derive that $p_5 = 0$. So effectively this becomes a two-mode two-state($\{|\tilde{\psi_i}\rangle\} = \{|0,1\rangle, |1,0\rangle\}$) protocol with the channel capacity of 1 *bit*. Therefore, under the condition of no vacuum detection, our entanglement-based protocol allows one to communicate one extra bit of information when compared to the best possible entanglement-free two-mode communication scheme.

Lastly, let us consider what happens when the vacuum detection is allowed but Bob's detectors are non-ideal.

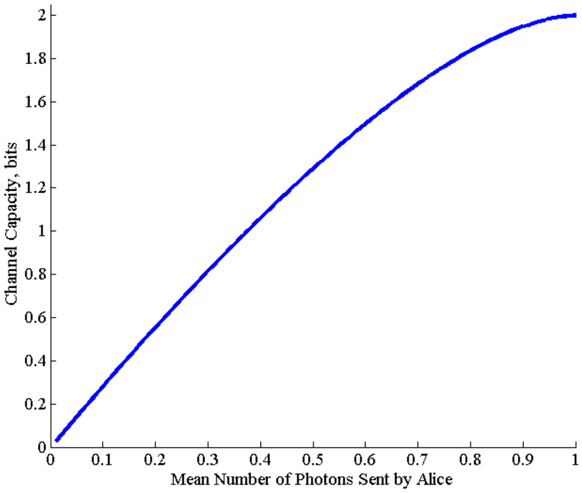

FIG. 3: (Color on-line) Channel capacity as a function of the mean number of photons communicated by Alice.

We show in Appendix A that in this case the difference in the channel capacity of the proposed entanglement-based protocol and its resource-equivalent entanglement-free analogue discussed earlier in this subsection is a positive function of the detection efficiency (see Fig.(5) for details). Thus, we show that for realistic detectors our entanglement-based protocol always allows transmission of more information then its entanglement-free version with the same amount of resources used.

C. Channel Capacity as a Function of Average Photon Number

Next, let us study how much information can be communicated in our entanglement-assisted protocol by sending less than one photon on average. This can be readily done by modifying the optimization constraint in Eq.(1). Note that, since Alice only uses passive optical elements, the average number of photons she sends to Bob is actually controlled by the source i.e.

$$\langle N_{Alice}\rangle = \langle\Psi_{in}|a^{\dagger}_{A1}a_{A1} + a^{\dagger}_{A2}a_{A2}|\Psi_{in}\rangle, \qquad (14)$$

where $a_{A1,A2}$ denote photon annihilation operators for Alice's modes $1, 2$. Thus, the new optimization problem at hand is,

$$\begin{aligned} &maximize\ I(\psi;\phi) \\ &s.t.\ \langle N_{Alice}\rangle = n, \end{aligned} \qquad (15)$$

where n is a constant $\in [0, 1]$. We used a gradient-based solver to optimize Eq.(15) numerically for various values of the average photon number n. For a fixed value of $n = 0.05$ we ran 500 independent optimizations using random starting points. We then selected the largest value of $\max I(\psi;\phi)$ over these runs and used it as a starting point for optimizing $I(\psi;\phi)$ for the next value of $n = 0.1$. By gradually varying the constraint value we calculated $\max I(\psi;\phi)$ as a function of n depicted in Fig.(3). We observe that if Alice sends Bob less then one photon on average then their channel capacity falls below two bits. However, they can achieve channel capacity $C \approx 1.63$ *bits* (the best value demonstrated to date [5]) by communicating $\approx 0.68 < 1$ photons on average.

D. Communicating Larger Alphabets

In principle, Alice may want to use larger alphabets than just the four-symbol one. After all, we are operating with the states in $\mathbb{C}^{10}$ and naturally the question arises whether a physical setup exists that attains the channel capacity $C = \log_2 M$ for some integer $M \in [5, 10]$. To answer this question we modified our protocol by allowing Alice to perform $M > 4$ unitary transformations on her two modes. At the same time we still require Bob to measure in the Fock basis $\{|\phi_j\rangle\}, j = 1, 10$. We numerically optimized the channel capacity in Eq.(2) for the cases of $M = 5, \cdots, 10$ and normalized the respective values to the maximal theoretically attainable channel capacity ($C = \log_2 M$). The results are plotted in Fig.(4). We notice that the maximal theoretical channel capacity is only achievable for the case of the two-bit alphabet (M=4). When Alice is trying to use $M > 4$ symbols in her alphabet the normalized channel capacity decreases. This is because Bob is unable to deterministically discriminate between the states $|\psi_i\rangle, i = 1, \cdots, M; M > 4$ by using projective measurements in Fock basis.

IV. OPTIMIZATION RESULTS FOR $N = 6, 8$ MODES $n = 2$ PHOTONS STATES

In principle, the protocol described in Sec. II can be used for linear optical circuits with arbitrary number of modes N and photons n. Here, we study two simple extensions with two photons ($n = 2$) distributed over $N = 6$ and $N = 8$ modes. We assume that in both cases Alice and Bob receive $N/2$ modes from the source. Our goal is again to solve numerically the channel capacity problem posed in Eq.(2).

For the $n = 2$, $N = 6$ case the dimensionality of the correspondent Hilbert space is $\dim \mathbb{C} = \frac{(n+N-1)!}{n!(N-1)!} = 21$ which naturally leads to the question: is it possible to design an entanglement-assisted communication scheme that provides channel capacity of $\log_2 21 \approx 4.39$ *bits* by sending just three modes from Alice to Bob? The necessary condition for that is Alice must be able to prepare 21 orthogonal states from an input state $|\Psi_{in}\rangle$ by means of "local" three-mode unitary transformations. However, we discovered numerically that Alice can at best prepare

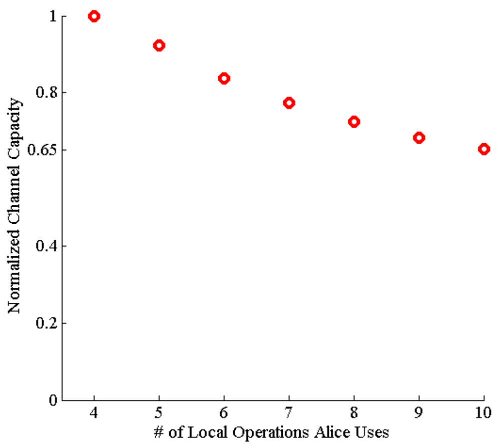

FIG. 4: (Color on-line) Normalized channel capacity. Red circles denote the channel capacity of our entanglement-assisted protocol obtained by solving Eq.(2) as a function of the number of Alice's local operations M normalized to the theoretical maximum of $\log_2 M$.

12 orthogonal states using three-mode unitary transformations. This means that the channel capacity cannot possibly exceed $\log_2 12 \approx 3.58$ *bits*. Next, we numerically solved the optimization task in a modified version of Eq.(2) for the case of $M = 12$ local $U(3)$ operations performed by Alice and one global $U(6)$ mode transformation performed by Bob. Obtained solutions suggest $C = 3.0$ *bits* which implies that even if Alice can locally prepare 12 orthogonal states, Bob cannot discriminate then deterministically by means of linear optics and coincidence detection. Indeed, we discovered that Bob can only perform a non-ambiguous detection of 8 orthogonal states encoded by Alice. Therefore, the practical(achievable) channel capacity in the case of $n = 2$ photons in $N = 6$ modes is limited to 3 *bits*. This result was obtained using an unconstrained optimization of the mutual information function. The actual mean number of photons per character in this scheme is 1.5 photons. Correspondent entanglement-free three-mode communication schemes with the same photon-per-character cost are equivalent to a six state protocol where Alice sends states $\{|1,0,0\rangle, |0,1,0\rangle, |0,0,1\rangle, |1,1,0\rangle|1,0,1\rangle, |0,1,1\rangle\}$ which results in the channel capacity of $\log_2 6 \approx 2.585$ *bits*(assuming perfect detectors and no vacuum detection constraint). Therefore, the information gain in the $n = 2$ photon, $N = 6$ modes dense coding scheme is $\approx 0.415 < 1$ *bits*.

Similar analysis for the case of $n = 2$ photons in $N = 8$ modes revealed that the practical channel capacity of the eight-mode communication is limited to $\log_2 12 \approx 3.58$ *bits* (compare to the four-mode entanglement-free channel capacity of $\log_2 10 \approx 2.3$ *bits*).

V. SUMMARY

We discussed the problem of entanglement-assisted communication channel capacity gain for a linear optical circuit with N modes populated by n photons. We discovered in the case of $N = 4$, $n = 2$ there is a class of mode-entangled states that supports protocols with a range of channel capacity gain over corresponding resource-equivalent entanglement-free protocols. We studied numerically 6 and 8 mode extensions of the protocol and provided estimates for the channel capacity in those cases.

Acknowledgments

This work was performed at Oak Ridge National Laboratory, operated by UT-Battelle for the U.S. Department of energy under contract no. DE-AC05-00OR22725. PL would like to thank Warren P. Grice for useful discussion. DBU acknowledges support from the NSF under grant PHY-1005709.

Appendix A: Channel Capacity with Imperfect Detectors

First, let us discuss why the detection of vacuum with a photon-counting detector is less efficient than the detection of a photon with the same detector. We assume that for any input state of light the detector produces either a "click" outcome (denoted as +) or a "no click" outcome (denoted as −). Next, we denote $p(-|0) = v$ the conditional probability of the detector producing a "no click" outcome, given the input state on the detector was vacuum $|0\rangle$. We also introduce the conditional probability $p(+|1) = s$ of the detector clicking provided the input state was a single photon $|1\rangle$. In a similar fashion we define the conditional probability $p(+|0) = 1 - v$ for the detector to click erroneously when the input state was $|0\rangle$ and the probability $p(-|1) = 1-s$ for the "no click" event when the input contained a photon. Note that for imperfect detectors "no click"("click") does not necessary imply that the input state was zero(one) photon.

Assuming that states $|0\rangle$ and $|1\rangle$ have equal probabilities of arriving at the detector i.e. $p(0) = p(1) = \frac{1}{2}$ let us calculate the probability $p(0|-)$ of the vacuum state arriving at the detector provided a "no click" event was recorded. Using Bayes' rule we obtain,

$$p(0|-) = \frac{p(-|0) \cdot p(0)}{p_{NC}} = \frac{v}{1 - s + v}, \qquad \text{(A1)}$$

where the total probability of a"no click" event p_{NC} is given by

$$p_{NC} = p(-|0) \cdot p(0) + p(-|1) \cdot p(1) = \frac{1 - s + v}{2}. \qquad \text{(A2)}$$

Similarly, we can compute the probability $p(1|+)$ of a single photon arriving at the detector provided a "click" event was recorded as

$$p(1|+) = \frac{p(C||0\rangle)\cdot p(|0\rangle)}{p_C} = \frac{s}{1+s-v}, \qquad \text{(A3)}$$

where the total probability of a "click" event p_C is given by

$$p_C = p(+|0)\cdot p(0) + p(+|1)\cdot p(1) = \frac{1+s-v}{2}. \qquad \text{(A4)}$$

Comparing Eq.(A1) and Eq.(A3) for a fixed value of v ($v \approx 1$) we observe that $p(1|+) > p(0|-)$ for any finite efficiency detector ($s < 1$). For a typical state-of-the-art superconducting single photon detector [11] $s \leq 0.9$ and $1 - v \approx 10^{-4}$ which translates into the vacuum detection efficiency of $p(0|-) \approx 0.91$. On the other, hand the efficiency of a single photon detection is $p(1|+) = 0.9999$. Therefore, detecting a single photon in a single mode is almost 10% more efficient than detecting vacuum. Furthermore, for the two-mode vacuum state $|0,0\rangle$ the detection efficiency is ≈ 0.82 which is significantly less than the detection efficiency for states $|1,1\rangle$ (≈ 0.9998) and $|0,1\rangle$ (≈ 0.91). Such a disparity in detection efficiency suggests that linear optical schemes that rely on double vacuum detection will experience a more severe channel capacity reduction than coincidence-based schemes.

With the previous discussion in mind, let us now quantify the effect of imperfect detectors onto the achievable channel capacity for both entanglement-free and entanglement-assisted communication protocols proposed in Section III. In the entanglement-free case, Bob uses two imperfect photon detectors to resolve between four possible states in Alice's alphabet $\mathcal{X} = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Therefore, Bob's alphabet contains four possible detection outcomes $\mathcal{Y} = \{--, -+, +-, ++\}$. Assuming that Bob's detectors have the same efficiency $s < 1, v \approx 1$ we can calculate the mutual information function $I(\mathcal{X};\mathcal{Y})$ between Alice and Bob as,

$$I(\mathcal{X};\mathcal{Y}) = \sum_{j=1}^{4}\sum_{k=1}^{4} p(x_j, y_k) \log \frac{p(x_j, y_k)}{p(x_j)p(y_k)}, \qquad \text{(A5)}$$

where the probability $p(x_j, y_k)$ can be expressed in the matrix form,

$$\begin{bmatrix} v^2 p_1 & v(1-v)p_1 & v(1-v)p_1 & (1-v)^2 p_1 \\ v(1-s)p_2 & vsp_2 & (1-s)(1-v)p_2 & s(1-v)p_2 \\ v(1-s)p_3 & (1-v)(1-s)p_3 & svp_3 & s(1-v)p_3 \\ (1-s)^2 p_4 & s(1-s)p_4 & s(1-s)p_4 & s^2 p_4 \end{bmatrix} \qquad \text{(A6)}$$

with the entry in the j-th row and k-th column representing a joint probability of Alice sending a state x_j from $\mathcal{X}$ and Bob detecting an outcome y_k from $\mathcal{Y}$. Here $p_i, i = 1,4$ is the probability of Alice sending the i-th state from her alphabet i.e. $p_i = p(x_i)\forall i$. An explicit expression for the mutual information can be readily obtained by combining Eqs.(A5-A6). For fixed values of s and v, $I(\mathcal{X};\mathcal{Y})$ is a function of only two independent parameters e.g. p_1 and p_2 (p_3 and p_4 can be expressed in terms of p_1 and p_2 using Eq.(12) and probability normalization) and maximizing it numerically gives the channel capacity C_{noent} for the entanglement-free communication scheme. In a similar fashion we can compute the channel capacity for the entanglement-assisted protocol. In this case Alice prepares one of the four states from the set $\mathcal{X} = \{|\psi_1\rangle, \cdots, |\psi_4\rangle\}$ defined in Section II and Bob registers one of 16 possible four-mode click patterns $\mathcal{Y} = \{- - --, + - --, \cdots, + + ++\}$. The mutual information $I(\mathcal{X};\mathcal{Y})$ can now be calculated by extending the summation range over k from 4 to 16 in Eq.(A5) and introducing a correspondent 4×16 join probability matrix $p(x_j, y_k)$. Note that unlike in the entanglement-free protocol now $I(\mathcal{X};\mathcal{Y})$ depends on three parameters $p_1, \cdots, p_3$ for a fixed value of s and v. The final algebraic expression can be optimized numerically to find the channel capacity of the entanglement-assisted protocol C_{ent}.

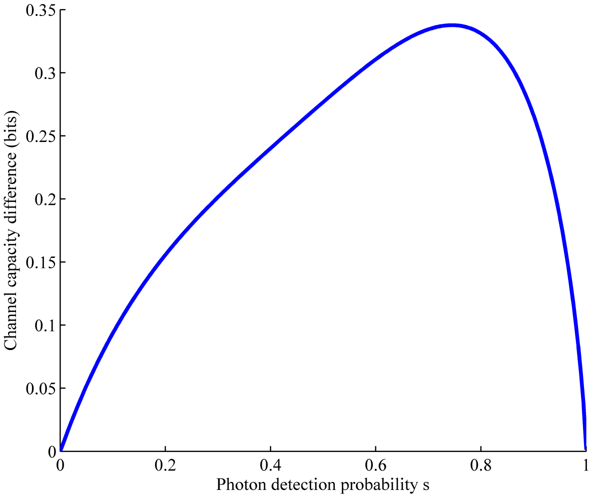

FIG. 5: (Color on-line) Channel capacity difference between entanglement-assisted and entanglement-free communication protocols as a function of the photon detection probability s.

We have obtained values of $\Delta C = C_{ent} - C_{noent}$ numerically by using a random search method on a set of 10^7 randomly generated probability distributions $\{p_i\}$ for various values of the single photon detection probability s and $v = 0.9999$. The results are displayed on Fig.(5). We noticed that in the limit of perfect detectors both protocols reach the same channel capacity (2 *bits*). However, with realistic detectors ($s < 1$) the entanglement-assisted protocol exhibits an information gain over its entanglement-free counterpart. In particular, for the best superconducting detectors ($s \approx 0.9$) the information gain is ≈ 0.27 *bits*. Although, the gain is < 1 *bit*, it is greater than the actual experimental gain of 0.13 *bits* observed in [2].

[1] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).

[2] K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger, Phys. Rev. Lett. **76**, 4656 (1996).

[3] C. Schuck, G. Huber, C. Kurtsiefer, and H. Weinfurter, Phys. Rev. Lett. **96**, 190501 (2006)

[4] M. Barbieri, G. Vallone, P. Mataloni, and F. De Martini, Phys. Rev. A **75**, 042317 (2007)

[5] J. T. Barreiro, T.-C. Wei and P. G. Kwiat, Nature Physics **4**, 282 (2008).

[6] L. Vaidman and N. Yoran, Phys. Rev. A **59**, 116 (1999); N. Lutkenhaus, J. Calsamiglia, and K. A. Suominen, Phys. Rev. A **59**, 3295 (1999).

[7] P. G. Kwiat and H. Weinfurter, Phys. Rev. A **58**, R2623 (1998).

[8] W. P. Grice, Phys. Rev. A **84**, 042331 (2011).

[9] F. Ewert and P. van Loock, Phys. Rev. Lett. **113**, 140403 (2014)

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley , 2006)

[11] F. Marsili *et al.,* Nature Photonics **7**, 210 (2013)

[12] only 9 of the coefficients c_k are independent because of the normalization constraint $\sum_{k=1}^{10} |c_k|^2 = 1$ and one of the phases can be set to 0